

KLASIFIKASI BERITA BERBAHASA INDONESIA MENGUNAKAN *NAÏVE BAYES CLASSIFIER*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

QURROTA 'AINI MUTHMAINNAH
NIM : 09021381520083

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

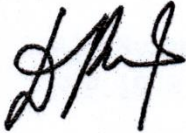
KLASIFIKASI BERITA BERBAHASA INDONESIA MENGUNAKAN *NAÏVE BAYES CLASSIFIER*

Oleh :

QURROTA 'AINI MUTHMAINNAH
NIM : 09021381520083

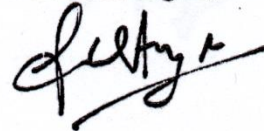
Palembang, Juli 2020

Pembimbing I,



Dian Palupi Rini, M.Kom., Ph.D
NIP. 197802232006042002

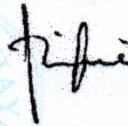
Pembimbing II,



Desty Rodiah, M.T.
NIP. -

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rifkie Primartha, MT
NIP. 197706012009121004

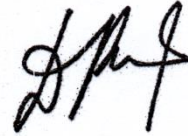


TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari tanggal Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

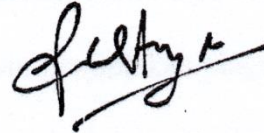
Nama : Qurrota 'Aini Muthmainnah
NIM : 09021381520083
Judul : Klasifikasi Berita Berbahasa Indonesia Menggunakan *Naïve Bayes Classifier*

1. Pembimbing I



Dian Palupi Rini, M.Kom., Ph.D
NIP. 197802232006042002

2. Pembimbing II



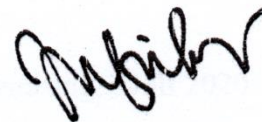
Desty Rodiah, M.T
NIP. -

3. Penguji I



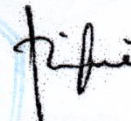
Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

4. Penguji II



Nabila Rizky Oktadini, S.Kom, M.T
NIP. 199110102018032001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, MT
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Qurrota 'Aini Muthmainnah
NIM : 09021381520083
Program Studi : Teknik Informatika
Judul : Klasifikasi Berita Berbahasa Indonesia
Menggunakan *Naïve Bayes Classifier*

Hasil Pengecekan Software
iThenticate/Turnitin : 19%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Juli 2020



Qurrota 'Aini Muthmainnah
NIM. 09021381520083

KATA PENGANTAR

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Tri Yulisman Eka Putra dan Rini Amirin, kakak dan adikku, Istiqlal Miftahul Jannah, Muhammad Abdurrahman Hidayatullah, Nurul Zukhrufa dan seluruh keluarga besarku yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika, dan Ibu Alvi Syahrini Utami, M.Kom selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku dosen pembimbing I dan Ibu Desty Rodiah, M.T selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Osvari Arsalan, M.T. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Ibu Alvi Syahrini Utami, M.Kom selaku dosen penguji I dan Ibu Nabila Rizky Oktadini, S.Kom, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Mbak Wiwin, dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Dr. Siti Hajar Othman selaku pembimbing saya selama menjalani perkuliahan di University Technology Malaysia.
9. Mbak Wiwin, dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Sahabatku Vivi Alvi Fatimah, teman seperjuangan Oktefvia Aruda Lisjana, Saniyah, Prilly Rizkytami, Limar Susilowati, Dhea Dwiputri, Elni Kurnia Sari, Ria Angellina dan Mayang Hermeiliza Eka Putri yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.
11. Teman-teman IFBIL A 2015 dan seluruh teman-teman Teknik Informatika Universitas Sriwijaya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Juli 2020

Qurrota 'Aini Muthmainnah

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR ISTILAH DAN SINGKATAN	xviii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Perumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 <i>Text Mining</i>	II-1

2.3	<i>Text Classification</i>	II-2
2.4	<i>Text Preprocessing</i>	II-2
2.4.1	<i>Case Folding</i>	II-3
2.4.2	<i>Tokenization</i>	II-3
2.4.3	<i>Stopword Removal</i>	II-3
2.4.4	<i>Stemming</i>	II-3
2.5	<i>Term Weighting</i>	II-4
2.6	<i>Naïve Bayes Classifier</i>	II-5
2.7	<i>Performance Measurement</i>	II-7
2.8	<i>Rational Unified Process (RUP)</i>	II-7
2.9	Penelitian Lain yang Relevan	II-10
2.10	Kesimpulan	II-11
BAB III METODOLOGI PENELITIAN		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1.	Jenis Data	III-1
3.2.2.	Sumber Data	III-1
3.2.3.	Metode Pengumpulan	III-2
3.3	Tahapan Penelitian	III-2
3.3.1.	Menetapkan Kerangka Kerja	III-3
3.3.2.	Menetapkan Kriteria Pengujian	III-10
3.3.3.	Menetapkan Format Data Pengujian	III-11
3.3.4.	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-12
3.3.5.	Melakukan Pengujian Penelitian	III-12
3.3.6.	Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-12
3.4	Metode Pengembangan Perangkat Lunak	III-14
3.4.1	Fase Insepsi	III-14

3.4.2	Fase Elaborasi	III-15
3.4.3	Fase Konstruksi	III-15
3.4.4	Fase Transisi	III-16
3.5	Manajemen Proyek Penelitian	III-16

BAB IV	PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1.	Pendahuluan	IV-1
4.2.	Rational Unified Process (RUP)	IV-1
4.2.1.	Fase Insepsi	IV-1
4.2.1.1.	Pemodelan Bisnis	IV-1
4.2.1.2.	Kebutuhan Sistem	IV-2
4.2.1.3.	Analisis dan Desain	IV-4
4.2.1.3.1.	Analisis Kebutuhan Perangkat Lunak.....	IV-4
4.2.1.3.2.	Analisis Data	IV-5
4.2.1.3.3.	Analisis Praproses Data	IV-6
4.2.1.3.4.	Analisis Naïve Bayes Classifier	IV-13
4.2.1.3.5.	Desain Perangkat Lunak	IV-20
4.2.2.	Fase Elaborasi	IV-28
4.2.2.1.	Pemodelan Bisnis	IV-28
4.2.2.1.1.	Perancangan Data	IV-28
4.2.2.1.2.	Perancangan Antarmuka	IV-29
4.2.2.2.	Kebutuhan Sistem	IV-30
4.2.2.3.	Diagram	IV-30
4.2.2.3.1.	Diagram Aktivitas	IV-30
4.2.2.3.2.	Diagram Sequence	IV-35
4.2.3.	Fase Konstruksi	IV-40
4.2.3.1.	Kebutuhan Sistem	IV-40
4.2.3.2.	Diagram Kelas	IV-40

4.2.3.3. Kelas Analisis	IV-42
4.2.3.4. Implementasi	IV-44
4.2.3.4.1. Implementasi Kelas	IV-44
4.2.3.4.2. Implementasi Antarmuka	IV-47
4.2.4. Fase Transisi	IV-48
4.2.4.1. Pemodelan Bisnis	IV-48
4.2.4.2. Kebutuhan Sistem	IV-49
4.2.4.3. Rencana Pengujian	IV-49
4.2.4.3.1. Rencana Pengujian Use Case Pilih File.....	IV-49
4.2.4.3.2. Rencana Pengujian Use Case Melakukan Praproses	IV-50
4.2.4.3.3. Rencana Pengujian Use Case Melakukan Klasifikasi	IV-50
4.2.4.3.4. Rencana Pengujian Use Case Menyimpan Hasil	IV-51
4.2.4.3.5. Rencana Pengujian Use Case Melihat Hasil Klasifikasi ...	IV-51
4.2.4.4. Implementasi	IV-51
4.2.4.4.1. Pengujian Use Case Pilih File	IV-52
4.2.4.4.2. Pengujian Use Case Melakukan Praproses	IV-53
4.2.4.4.3. Pengujian Use Case Melakukan Klasifikasi	IV-53
4.2.4.4.4. Pengujian Use Case Menyimpan Hasil Klasifikasi	IV-54
4.2.4.4.5. Pengujian Use Case Melihat Hasil Klasifikasi	IV-55
4.3. Kesimpulan	IV-56
BAB V HASIL DAN ANALISIS PENELITIAN	V-1
5.1. Pendahuluan	V-1
5.2. Data Hasil Percobaan/Penelitian	V-1
5.2.1. Konfigurasi Percobaan	V-1
5.2.2. Hasil Pengujian	V-2
5.3. Analisis Hasil Penelitian	V-6
5.4. Kesimpulan	V-7

BAB VI KESIMPULAN DAN SARAN	VI-1
6.1. Pendahuluan	VI-1
6.2. Kesimpulan	VI-1
6.3. Saran	VI-2
 DAFTAR PUSTAKA	 xx
LAMPIRAN	xxii

DAFTAR TABEL

	Halaman
III-1. Teks Berita	III-5
III-2. <i>Case Folding</i>	III-5
III-3. <i>Tokenization</i>	III-6
III-4. <i>Stopwords Removal</i>	III-7
III-5. <i>Stemming</i>	III-8
III-6. <i>Term Frequency</i>	III-9
III-7. Pembagian Data untuk Pengujian	III-11
III-8. Rancangan Tabel Hasil Uji Klasifikasi	III-11
III-9. Rancangan Tabel Hasil Analisis Klasifikasi	III-13
III-10. Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i>	III-18
IV-1. Kebutuhan Fungsional	IV-3
IV-2. Kebutuhan Non-Fungsional	IV-4
IV-3. Teks Berita	IV-5
IV-4. <i>Case Folding</i>	IV-6
IV-5. <i>Tokenization</i>	IV-7
IV-6. <i>Stopword Removal</i>	IV-8
IV-7. <i>Stemming</i>	IV-11
IV-8. <i>Term Weighting</i>	IV-13
IV-9. Contoh Data Latih	IV-14
IV-10. Contoh Perhitungan <i>Prior</i> dan <i>Conditional Probability</i>	IV-18
IV-11. Definisi Aktor	IV-21
IV-12. Definisi Use Case	IV-22

IV-13. Skenario <i>Use Case</i> Pilih File	IV-23
IV-14. Skenario <i>Use Case</i> Melakukan Praproses	IV-24
IV-15. Skenario <i>Use Case</i> Melakukan Klasifikasi	IV-25
IV-16. Skenario <i>Use Case</i> Menyimpan Hasil Klasifikasi	IV-26
IV-17. Skenario <i>Use Case</i> Menampilkan Hasil Klasifikasi	IV-27
IV-18. Implementasi Kelas	IV-44
IV-19. Rencana Pengujian <i>Use Case</i> Pilih File	IV-49
IV-20. Rencana Pengujian <i>Use Case</i> Melakukan Praproses	IV-50
IV-21. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi	IV-50
IV-22. Rencana Pengujian <i>Use Case</i> Menyimpan Hasil Klasifikasi	IV-51
IV-23. Rencana Pengujian <i>Use Case</i> Menampilkan Hasil Klasifikasi	IV-51
IV-24. Pengujian <i>Use Case</i> Pilih File	IV-52
IV-25. Pengujian <i>Use Case</i> Melakukan Praproses	IV-53
IV-26. Pengujian <i>Use Case</i> Melakukan Klasifikasi	IV-54
IV-27. Pengujian <i>Use Case</i> Menyimpan Hasil Klasifikasi	IV-54
IV-28. Pengujian <i>Use Case</i> Menampilkan Hasil Klasifikasi	IV-55
V-1. Tabel Hasil Pengujian Perangkat Lunak pada Data Uji	V-2
V-2. Hasil Akurasi Pengujian Pertama	V-4
V-3. Hasil Akurasi Pengujian Kedua	V-4
V-4. Hasil Akurasi Pengujian Ketiga	V-5
V-5. Hasil Akurasi Pengujian Keempat	V-5

DAFTAR GAMBAR

	Halaman
II-1. Arsitektur proses RUP (Kruchten, 2001)	II-8
III-1. Diagram Tahapan Penelitian	III-2
III-2. Diagram Tahapan Proses Perangkat Lunak	III-4
III-3. Tahapan Pengujian Penelitian	III-13
III-4. Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-26
III-5. Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-26
III-6. Penjadwalan untuk Tahap Menentukan Kriteria Pengujian	III-27
III-7. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi	III-27
III-8. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaborasi	III-28
III-9. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi	III-28
III-10. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi	III-29
III-11. Penjadwalan untuk Tahap Melakukan Pengujian Penelitian	III-29
III-12. Penjadwalan untuk Tahap Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan	III-30
IV-1. Diagram <i>Use Case</i>	IV-21
IV-2. Perancangan Antarmuka Perangkat Lunak	IV-29
IV-3. Diagram Aktivitas Pilih <i>File</i>	IV-31
IV-4. Diagram Aktivitas Melakukan Praproses	IV-32
IV-5. Diagram Aktivitas Melakukan Klasifikasi	IV-33

IV-6.	Diagram Aktivitas Menyimpan Hasil Klasifikasi	IV-34
IV-7.	Diagram Aktivitas Menampilkan Hasil Klasifikasi	IV-34
IV-8.	Diagram <i>Sequence</i> Pilih <i>File</i>	IV-35
IV-9.	Diagram <i>Sequence</i> Melakukan Praproses	IV-36
IV-10.	Diagram <i>Sequence</i> Melakukan Klasifikasi	IV-38
IV-11.	Diagram <i>Sequence</i> Menyimpan Hasil Klasifikasi	IV-39
IV-12.	Diagram <i>Sequence</i> Menampilkan Hasil Klasifikasi	IV-39
IV-13.	Diagram Kelas	IV-41
IV-14.	Kelas Analisis <i>Use Case</i> Pilih File	IV-42
IV-15.	Kelas Analisis <i>Use Case</i> Melakukan Praproses	IV-42
IV-16.	Kelas Analisis <i>Use Case</i> Melakukan Klasifikasi	IV-43
IV-17.	Kelas Analisis <i>Use Case</i> Menyimpan Hasil Klasifikasi	IV-43
IV-18.	Kelas Analisis <i>Use Case</i> Menampilkan Hasil Klasifikasi	IV-44
IV-19.	Implementasi Antarmuka Perangkat Lunak	IV-48
V-1.	Perbandingan Nilai Akurasi Klasifikasi Teks Berita	V-6

DAFTAR LAMPIRAN

Halaman

1. Pengujian Perangkat Lunak L-1
2. Kode Program L-11

INDONESIAN NEWS CLASSIFICATION USING NAÏVE BAYES CLASSIFIER

By:

Qurrota 'Aini Muthmainnah
09021381520083

ABSTRACT

News was initially published through media such as television, radio and newspapers, but with the current technological advancements making a digitizing information gets easier. News in a form of digital text can be published faster, actual and inexpensive which enable it to increased gradually. Therefore, there is a need for a system that can classify news automatically according to existing news categories by using the text classification method. Thus a very large collection of documents can be organized in order to simplify and speed up the search for information related. In this research, the news text classification used were Naïve Bayes Classifier method which able to classify into four categories namely, natural disasters, health, sports and education. The test is carried out four times with different data sharing, and the accuracy obtained is the first test 100%, the second test 100%, the third test 98.33% and the fourth test 96.25%. From these results, it can be concluded that the results of news text classification are good.

Keywords : Text Classification, Naïve Bayes Classifier.

KLASIFIKASI BERITA BERBAHASA INDONESIA
MENGUNAKAN *NAÏVE BAYES CLASSIFIER*

Oleh:

Qurrota 'Aini Muthmainnah
09021381520083

ABSTRAK

Berita pada awalnya disalurkan melalui media seperti televisi, radio dan koran, namun dengan kemajuan teknologi saat ini membuat digitalisasi informasi lebih mudah, berita berbentuk teks digital lebih cepat tersebar, aktual dan murah, sehingga dapat mengalami pelonjakan yang besar. Oleh karena itu, perlu adanya sistem yang bisa mengklasifikasikan berita secara otomatis sesuai dengan kategori-kategori berita yang ada, dengan menggunakan metode klasifikasi teks, maka kumpulan dokumen yang jumlahnya sangat besar tersebut dapat diorganisir, sehingga dapat mempermudah dan mempercepat pencarian informasi yang dibutuhkan. Dalam penelitian ini, klasifikasi teks berita menggunakan metode *Naive Bayes Classifier* untuk mengklasifikasikan ke dalam empat kategori yaitu, bencana alam, kesehatan, olahraga dan pendidikan. Pengujian dilakukan sebanyak empat kali dengan pembagian data yang berbeda-beda, dan hasil akurasi yang didapat yaitu pengujian pertama 100%, pengujian kedua 100%, pengujian ketiga 98,33% dan pengujian keempat 96,25%. Dari hasil tersebut, dapat disimpulkan bahwa hasil klasifikasi teks berita sudah baik.

Kata Kunci : Klasifikasi Teks, *Naive Bayes Classifier* (NBC)

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini membahas mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dalam laporan ini dan terakhir kesimpulan.

1.2 Latar Belakang

Berita adalah informasi mengenai peristiwa yang ingin diketahui oleh publik, bersifat aktual, terjadi di lingkungan pembaca, yang melibatkan tokoh-tokoh terkemuka, yang akibat peristiwa tersebut dapat berpengaruh terhadap pembaca (Nasution, 2008). Sebagai makhluk sosial, manusia akan selalu membutuhkan berita atau informasi baru untuk menambah wawasannya dan untuk mendewasakan alam berpikirnya (Ermanto, 2005).

Dalam perkembangannya, pada awalnya berita disalurkan melalui media seperti televisi, radio, koran, dan sekarang dengan kehadiran *personal computer* (PC), *smartphone* dan perkembangan internet membuat digitalisasi informasi lebih mudah. Berita dapat tersebar lebih cepat, aktual dan murah yang berbentuk teks digital, sehingga dapat mengalami pelonjakan yang besar.

Banyaknya teks berita membuat pembaca sulit untuk menemukan jenis kategori yang diperlukan. Untuk mendapatkan jenis kategori dalam dokumen teks secara manual, pembaca harus membaca isi dokumen. Jika dokumen teks sangat

panjang, maka dibutuhkan waktu yang lama bagi pembaca untuk mendapatkan jenis kategori dari sebuah berita.

Oleh karena itu, perlu adanya sistem yang bisa mengklasifikasikan berita secara otomatis sesuai dengan kategori-kategori berita yang ada, dengan menggunakan metode klasifikasi teks, maka kumpulan dokumen yang jumlahnya sangat besar tersebut dapat diorganisir, sehingga dapat mempermudah dan mempercepat pencarian jenis kategori dari sebuah berita.

Text mining merupakan salah satu variasi teknik dari data mining yang dapat digunakan untuk menangani masalah *clustering*, *information extraction*, *information retrieval* dan *text classification* (klasifikasi teks), yang cara kerjanya yaitu menemukan pola menarik dari sekumpulan data yang besar. (Meaney et al., 2016). Klasifikasi adalah suatu proses menemukan model yang dapat membedakan kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek (Usyal, 2014).

Metode yang dapat digunakan dalam penelitian klasifikasi teks sangat bervariasi, salah satunya yaitu metode *Naïve Bayes Classifier*. *Naive Bayes Classifier* adalah sebuah algoritma analisa statistik, yang bekerja dengan mengolah data numerik, sering digunakan dalam penelitian tentang klasifikasi teks karena kesederhanaan dan efektivitasnya yang menggunakan ide dasar probabilitas gabungan dari kata-kata dan kategori untuk memperkirakan probabilitas kategori pada suatu dokumen (Chy et al., 2014).

Menurut (Socrates et al., 2016) kelebihan dari metode *Naive Bayes Classifier* adalah algoritma yang sederhana dengan kompleksitas perhitungan yang rendah. Algoritma ini memiliki akurasi yang tinggi (Rish, 2001).

Penelitian mengenai klasifikasi teks menggunakan metode *Naive Bayes Classifier* sudah pernah dilakukan oleh (Devita et al., 2018) yang mampu mengklasifikasikan artikel teks kedalam empat kategori yaitu pendidikan ekonomi, pendidikan bisnis manajemen, akuntansi aktual dan ekonomi bisnis. Praproses yang dipakai dalam penelitian ini yaitu *tokenization*, *case folding* dan *term weighting* (TF-IDF). Penelitian tersebut menghasilkan akurasi yang cukup baik, hal ini karena keunggulan dari metode *naive bayes* sendiri yaitu mampu melakukan klasifikasi meskipun memiliki data *training* yang sedikit untuk estimasi parameternya. Tingkat akurasi yang didapatkan sebesar 70%, dari 40 data yang diuji didapatkan 28 dokumen yang berhasil diklasifikasikan dengan benar oleh sistem.

Penelitian klasifikasi lainnya oleh (Fajar, 2013) yang mengklasifikasikan posting twitter kemacetan lalu lintas kota bandung, dari hasil pengujian akurasi sistem yang dikembangkan, menghasilkan nilai akurasi terkecil sebesar 78% pada proses pengujian dengan menggunakan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 91,60% pada proses pengujian dengan menggunakan sampel sebanyak 13106.

Berdasarkan uraian diatas, penelitian ini akan menerapkan metode *Naive Bayes Classifier* untuk klasifikasi teks berita ke dalam kategori-kategori yang ada, untuk membantu mempermudah dan mempercepat pengguna dalam pencarian informasi yang dibutuhkan.

1.3 Perumusan Masalah

Berdasarkan uraian latar belakang masalah, rumusan masalah dari penelitian ini adalah :

1. Bagaimana penerapan *Naive Bayes Classifier* untuk klasifikasi teks berita?
2. Bagaimana hasil tingkat akurasi klasifikasi teks berita menggunakan *Naive Bayes Classifier*?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengetahui cara kerja metode *Naive Bayes Classification* untuk klasifikasi teks berita berbahasa Indonesia.
2. Membangun model *Naive Bayes Classification* untuk digunakan dalam proses klasifikasi pada teks berita.
3. Mengetahui hasil tingkat akurasi klasifikasi teks berita dengan metode *Naive Bayes Classifier*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Hasil penelitian dapat digunakan untuk mempermudah klasifikasi dokumen berita dengan memperoleh informasi dengan cepat tanpa harus membacanya satu per satu dan dapat menemukan kategori berita secara otomatis.
2. Membantu penelitian klasifikasi dokumen berita yang mendatang sebagai referensi penelitian.

1.6 Batasan Masalah

Batasan masalah penelitian ini adalah :

1. Kategori dokumen berita yang akan diklasifikasikan bencana alam, kesehatan, olahraga dan pendidikan.
2. Jenis format teks berita yang akan digunakan adalah *.txt.
3. Jenis data yang digunakan adalah dokumen teks berita dalam Bahasa Indonesia.

1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir adalah sebagai berikut :

BAB I. PENDAHULUAN

Bab ini membahas mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dalam laporan ini.

BAB II. KAJIAN LITERATUR

Bab ini akan memaparkan dasar-dasar teori yang digunakan dalam penelitian, seperti penjelasan mengenai *text mining*, *text classification*, *Naïve Bayes Classifier*, serta penjelasan lainnya yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas mengenai tahapan penelitian yang akan dilaksanakan pada penelitian. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu

pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas tentang proses pengembangan perangkat lunak berdasarkan metode pengembangan perangkat lunak yang digunakan pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil dari pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian-uraian pada semua bab dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan penelitian selanjutnya.

1.8 Kesimpulan

Penelitian yang akan dilakukan adalah pengklasifikasian teks berita berbahasa Indonesia dengan menggunakan metode *Naïve Bayes Classifier*. Penelitian ini ditujukan untuk mengetahui cara kerja dan melihat hasil akurasi dari metode *Naïve Bayes Classifier* dalam mengklasifikasikan teks berita yang dibagi kedalam beberapa kelas atau kategori, yaitu bencana alam, kesehatan, olahraga dan pendidikan.

DAFTAR PUSTAKA

- Aldrich, J. (2008). "RA Fisher on Bayes and Bayes' theorem." *Bayesian Analysis* 3(1): 161-170.
- Chandra, D. N., Indrawan, G., & Sukajaya, I. N. (2016). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmiah Teknologi Informasi Asia*, 10(1), 11–19.
- Chy, A. N., Seddiqui, H., & Das, S. (2014). Bangla news classification using naive Bayes classifier. In *16th Int'l Conf. Computer and Information Technology, ICCIT 2013*. <https://doi.org/10.1109/ICCITechn.2014.6997369>
- Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427. <https://doi.org/10.25126/jtiik.201854773>
- Domingos, P. and M. Pazzani (1997). "On the optimality of the simple Bayesian classifier under zero-one loss." *Machine learning* 29(2-3): 103-130.
- Ermanto. (2005). "Menjadi wartawan Handal dan Profesional". Yogyakarta: CintaPena.
- Fajar, R. S. E. W. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 7(1), 13–22. <https://doi.org/10.22146/ijccs.3048>
- Hearst, M. (2003). "What is text mining." SIMS, UC Berkeley.

- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 306–312.
- Indriyono, B. V., Utami, E., & Sunyoto, A. (2015). Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode Naïve Bayes Classifier (Studi Kasus: STMIK Kadiri). *Sisfo*, 05(04). <https://doi.org/10.24089/j.sisfo.2015.09.010>
- Meaney, C., Moineddin, R., Voruganti, T., O'Brien, M. A., Krueger, P., & Sullivan, F. (2016). Text mining describes the use of statistical and epidemiological methods in published medical research. *Journal of Clinical Epidemiology*, 74(December 2015), 124–132. <https://doi.org/10.1016/j.jclinepi.2015.10.020>
- Nafalski, A. and A. P. Wibawa (2016). "Machine translation with Javanese speech levels' classification." *Informatyka, Automatyka, Pomiarzy w Gospodarce i Ochronie Środowiska*.
- Palaniappan, S. and R. Awang (2008). Intelligent heart disease prediction system using data mining techniques. *Computer Systems and Applications*, 2008. AICCSA 2008. IEEE/ACS International Conference on, IEEE.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, IBM New York.
- Socrates, I. G. A., Akbar, A. L., & Akbar, M. S. (2016). *Optimasi Naïve Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio*. 7(1), 22–30. <https://doi.org/10.24843/LKJITI.2016.v07.i01.p0>